Lösningsförslag till
Matematisk statistik och sannolikhetslära
2020-06-02

1. Låt L, S och D beteckna händelserna "skruven är liten", "skruven är stor" respektive "skruven är defekt". Med tolkningen "likformig fördelning över paketen" är $P(L) = 2/3$, $P(S) = 1/3$, $P(D \mid L) = \frac{12}{200} = 0.06$ och $P(D \mid S) = \frac{8}{100} = 0.08$.

   (a) Lagen om total sannolikhet ger

   $$P(D) = P(L)P(D \mid L) + P(S)P(D \mid S) = \tfrac{2}{3} \cdot 0.06 + \tfrac{1}{3} \cdot 0.08 = \frac{1}{15} \approx 0.067.$$

   (c) Bayes sats ger

   $$P(S \mid D) = \frac{P(S)P(D \mid S)}{P(D)} = \frac{\frac{1}{3} \cdot 0.08}{\frac{1}{15}} = 0.4.$$

   (b) Vi har

   $$\frac{P(L \cap D)}{P(D)} = P(L \mid D) = 1 - P(S \mid D) = 0.6,$$

   så

   $$P(L \cap D) = P(D) \cdot 0.6 = \tfrac{1}{15} \cdot 0.6 = 0.04.$$

   Med tolkningen "likformig fördelning över skruvarna", som också är rimlig, fås $P(L) = 0.8$, $P(S) = 0.2$, $P(D) = 0.064$, $P(S \mid D) = 0.25$ och $P(L \cap D) = 0.048$.

   Svar: (a) 6.7 %, (b) 4 %, (c) 40 %, **eller** (a) 6.4 %, (b) 4.8 %, (c) 25 %.

2. (a) Vi har

   $$\begin{aligned} 1 &= \iint_{\mathbf{R}^2} f_{X,Y}(x,y)\,dxdy = c\int_0^2 \int_0^2 (x+y)\,dx\,dy = c\int_0^2 \left[\frac{x^2}{2} + xy\right]_{x=0}^2 dy \\ &= c\int_0^2 (2+2y)\,dy = c\Big[2y + y^2\Big]_0^2 = 8c, \end{aligned}$$

   så $c = 1/8$.

   (b) För $0 \le x \le 2$ gäller

   $$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \tfrac{1}{8}\int_0^2 (x+y)\,dy = \tfrac{1}{8}\left[xy + \frac{y^2}{2}\right]_{y=0}^2 = \frac{x+1}{4}.$$

   Vidare är

   $$EX = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^2 x \cdot \frac{x+1}{4}\,dx = \tfrac{1}{4}\int_0^2 (x^2 + x)\,dx = \tfrac{1}{4}\left[\frac{x^3}{3} + \frac{x^2}{2}\right]_0^2 = \frac{7}{6}.$$

(c) Rita figur! Vi har

$$\begin{aligned}
P(|X-Y|>1) &= P(Y<X-1)+P(Y>X+1) \\
&= \tfrac{1}{8}\int_0^1\int_{y+1}^2 (x+y)\,dx\,dy + \tfrac{1}{8}\int_0^1\int_{x+1}^2 (x+y)\,dy\,dx \\
&= 2\cdot\tfrac{1}{8}\int_0^1\int_{y+1}^2 (x+y)\,dx\,dy = \tfrac{1}{4}\int_0^1 \left[\frac{x^2}{2}+xy\right]_{x=y+1}^2 dy \\
&= \tfrac{1}{4}\int_0^1 \left(-\frac{3y^2}{2}+\frac{3}{2}\right) dy = \frac{1}{4}.
\end{aligned}$$

Svar: (a) $c = 1/8$, (b) $f_X(x) = (x+1)/4$, $0 \le x \le 2$; $EX = 7/6$, (c) $P(|X-Y|>1) = 1/4$.

3. Sats 3.4 ger

$$g(t) = Et^X = \sum_{k=0}^{n} t^k p_X(k),$$

så

$$g'(t) = \sum_{k=0}^{n} kt^{k-1}p_X(k) \tag{1}$$

och

$$g''(t) = \sum_{k=0}^{n} k(k-1)t^{k-2}p_X(k). \tag{2}$$

Sätt $t = 1$ i (1) och (2), så fås

$$g'(1) = \sum_{k=0}^{n} kp_X(k) = EX$$

respektive

$$g''(1) = \sum_{k=0}^{n} k(k-1)p_X(k) = \sum_{k=0}^{n} k^2 p_X(k) - \sum_{k=0}^{n} kp_X(k) = EX^2 - EX,$$

så $EX^2 = g''(1) + EX$, så

$$VX = EX^2 - (EX)^2 = g''(1) + g'(1) - (g'(1))^2.$$

4. (a) Låt X beteckna antalet tentander som får betyg 5. Då är $X \in \mathrm{Bin}(60, 0.06)$, så

$$P(X=k) = \binom{60}{k}\cdot 0.06^k \cdot 0.94^{60-k}.$$

Vi får

$$\begin{aligned}
P(X\ge 3) &= 1-P(X\le 2) = 1-\Big(P(X=0)+P(X=1)+P(X=2)\Big) \\
&= 1-\left(0.94^{60}+60\cdot 0.06\cdot 0.94^{59}+\binom{60}{2}\cdot 0.06^2\cdot 0.94^{58}\right) \approx 0.7060.
\end{aligned}$$

(b) Vi har $0.06 < 0.1$, så $X \approx \mathrm{Po}(3.6)$ är en lämplig approximation. Tabell 3 ger

$$P(X \geq 6) = 1 - P(X \leq 5) \approx 1 - 0.8441 = 0.1559.$$

(c) Beteckna antalet tentander med n. Sätt

$$Y_i = \begin{cases} 1 & \text{om tentand } i \text{ tar väska } i, \\ 0 & \text{annars.} \end{cases}$$

Då är $Y_i \in \mathrm{Be}(1/n)$, så $EY_i = 1/n$ och

$$VY_i = \frac{1}{n}\left(1 - \frac{1}{n}\right) = \frac{n-1}{n^2}.$$

Sätt $Y = Y_1 + \cdots + Y_n$. Vi söker EY och VY.
Vi har

$$EY = \sum_{i=1}^{n} EY_i = n \cdot \frac{1}{n} = 1.$$

Anta att $i \neq j$. Sats 3.28 ger

$$\begin{aligned} E(Y_iY_j) &= \sum_{k,l} kl \cdot P(Y_i = k, Y_j = l) = P(Y_i = 1, Y_j = 1) = P(Y_j = 1)P(Y_i = 1 \,|\, Y_j = 1) \\ &= \frac{1}{n} \cdot \frac{1}{n-1} = \frac{1}{n(n-1)}, \end{aligned}$$

så

$$C(Y_i, Y_j) = E(Y_iY_j) - EY_i \cdot EY_j = \frac{1}{n(n-1)} - \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n(n-1)} - \frac{1}{n^2},$$

så Sats 3.34 ger

$$\begin{aligned} VY &= \sum_{i=1}^{n} VY_i + \sum_{i \neq j} C(Y_i, Y_j) = n \cdot \frac{n-1}{n^2} + n(n-1) \cdot \left(\frac{1}{n(n-1)} - \frac{1}{n^2}\right) \\ &= \frac{n-1}{n} + 1 - \frac{n-1}{n} = 1. \end{aligned}$$

Svar: (a) 70.6 %, (b) 15.6 %, (c) både väntevärdet och variansen är 1.

5. (a) Låt X och Y beteckna livslängden för ett lysrör respektive antalet lysrör som fungerar efter två dagar. Då är $X \in \mathrm{Exp}(\lambda)$ och $Y \in \mathrm{Bin}(20, p)$, där p är sannolikheten att ett lysrör fungerar efter två dagar. Vi har

$$p = P(X \geq 2) = \int_2^{\infty} \lambda e^{-\lambda x}\, dx = \Big[-e^{-\lambda x} \Big]_2^{\infty} = e^{-2\lambda},$$

så $Y \in \mathrm{Bin}(20, e^{-2\lambda})$, så

$$p_Y(k) = \binom{20}{k} (e^{-2\lambda})^k (1 - e^{-2\lambda})^{20-k}.$$

Således är likelihoodfunktionen[1]

$$L(\lambda) = \binom{20}{17} e^{-34\lambda}(1 - e^{-2\lambda})^3,$$

så

$$\ln L(\lambda) = \ln \binom{20}{17} - 34\lambda + 3(1 - e^{-2\lambda}),$$

så

$$\frac{d}{d\lambda}\Big(\ln L(\lambda)\Big) = -34 + \frac{3}{1 - e^{-2\lambda}} \cdot D(1 - e^{-2\lambda}) = -34 + \frac{6e^{-2\lambda}}{1 - e^{-2\lambda}} = \frac{40e^{-2\lambda} - 34}{1 - e^{-2\lambda}},$$

så

$$\frac{d}{d\lambda}\Big(\ln L(\lambda)\Big) = 0 \iff e^{-2\lambda} = \frac{34}{40} \iff \lambda = -\tfrac{1}{2} \ln \frac{34}{40} = \tfrac{1}{2} \ln \frac{20}{17}.$$

Derivatans teckenväxling är $+0-$, så vi har maximum. Alltså är ML-skattningen $\lambda^* = \frac{1}{2} \ln \frac{20}{17} \approx 0.0813$. (Den förväntade livslängden för ett lysrör skattas alltså med $1/\lambda^* \approx 12.3$ dagar.)

(b) Låt X beteckna antalet kontaktade studenter som hade velat ha fler sådana uppgifter. Då är $X \in \mathrm{Hyp}(N, n, p)$, där $N = 800$ och $n = 150$, och $x = 85$, så $p^* = x/n \approx 0.57$. Vi ska testa H_0: $p = 0.5$ mot H_1: $p > 0.5$. Anta att H_0 gäller. Då är $VX = np(1-p)\frac{N-n}{N-1} \approx 31 > 5$, så CGS medför att $X \approx \mathrm{N}(np, VX)$, så $p^*(X) = X/n \approx \mathrm{N}(p, VX/n^2)$. Som testvariabel väljer vi därför

$$T = T(x) = \frac{x/n - p}{\sqrt{VX/n^2}},$$

som är en observation av $T(X) \approx \mathrm{N}(0, 1)$.

Med H_1: $p > 0.5$ blir det kritiska området $C = \{T \geq \lambda_{0.05}\}$. Vi får $T \approx 1.81 \geq 1.64 \approx \lambda_{0.05}$, så H_0 förkastas till förmån för H_1, dvs. lärarna har fått ett signifikant resultat vid 5 % felrisk.

Alternativ lösning: Den undre konfidensgränsen är

$$p^* - 1.6449\sqrt{VX/n^2} \approx p^* - 1.6449\sqrt{\frac{p^*(1-p^*)}{n}}\sqrt{\frac{N-n}{N-1}} \approx 0.5066;$$

$0.5 < 0.5066$, så H_0 förkastas till förmån för H_1.

Nödvändiga antaganden: De kontaktade studenterna valdes slumpmässigt. De kontaktade studenterna som inte gav något svar alls, hade inte velat ha fler sådana uppgifter; det är dock rimligt att anta att det hade stått i texten om det fanns icke-svarande studenter.

Svar: (a) $\lambda^* = \frac{1}{2} \ln \frac{20}{17} \approx 0.0813$, (b) ja, det har de.

[1] Räkningarna blir lite enklare om man istället betraktar $L(p) = \binom{20}{17} p^{17}(1-p)^3$, vilket ger $p^* = \frac{17}{20}$ och samma λ^* som ovan.

6. (a) Låt X och Y beteckna antalet procent av alla virus som inte dödas av 70- respektive 80-procentspriten. Då är

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{s_p\sqrt{1/4 + 1/5}} \in t(4+5-2) = t(7),$$

där

$$s_p = \sqrt{\frac{S_{xx} + S_{yy}}{4+5-2}} \approx 0.6067,$$

ty

$$S_{xx} = \sum_{i=1}^{4} (x_i - \bar{x})^2 = \sum_{i=1}^{4} x_i^2 - 4\bar{x}^2 \approx 1.1115$$

och

$$S_{yy} = \sum_{i=1}^{5} (y_i - \bar{y})^2 = \sum_{i=1}^{5} y_i^2 - 5\bar{y}^2 \approx 1.4654,$$

vilket ger konfidensintervallet

$$I_{\mu_1-\mu_2} = \left(\bar{x} - \bar{y} \pm t_{0.005}(7) \cdot s_p\sqrt{1/4 + 1/5}\right) \approx (-0.62, 2.23).$$

(b) Vi ska testa H_0: $\mu_1 = \mu_2$ mot H_1: $\mu_1 > \mu_2$. Som testvariabel väljer vi

$$T = T(x, y) = \frac{\bar{x} - \bar{y}}{s_p\sqrt{1/4 + 1/5}},$$

som är en observation av $T(X, Y) \in t(7)$ om H_0 gäller. Med H_1: $\mu_1 > \mu_2$ blir det kritiska området $C = \{T \geq t_{0.05}(7)\}$. Vi får $T \approx 1.98 \geq 1.89 \approx t_{0.05}(7)$, så H_0 förkastas till förmån för H_1, dvs. 80-procentspriten är bättre vid 5 % felrisk.

Alternativ lösning: Det ensidiga konfidensintervallet är $(0.03, 100)$, som inte innehåller 0, så H_0 förkastas till förmån för H_1.

Svar: (a) $I_{\mu_1-\mu_2} \approx (-0.62, 2.23)$, (b) ja, det kan man.